DOCUMENT RESUME

ED 317 713 CE 054 129

AUTHOR Horrigan, Michael W.
TITLE Data Needs for Labor Market Analysis. Background Paper No. 44.
SPONS AGENCY Department of Labor, Washington, DC. Commission on Workforce Quality and Labor Market Efficiency.
PUB DATE Sep 89
NOTE 29p.; In "Investing in People: A Strategy to Address America's Workforce Crisis" (CE 054 080).
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Data Analysis; *Data Collection; *Economic Research; *Employment Statistics; *Labor Market; Labor Needs; Labor Supply; Labor Turnover; Statistical Analysis; Surveys; Unemployment Insurance
IDENTIFIERS *Bureau of Labor Statistics; Current Population Survey

ABSTRACT

Opinions on how the Bureau of Labor Statistics (BLS) can better meet the data needs of users of government-provided labor market data were sought from users inside and outside government. The following recommendations, among others, are based on those opinions: (1) create a quick-response household survey capability at the BLS, using random digit dialing supplemented by a purchase list for households without telephones or a list area frame; (2) start a new youth cohort for the National Longitudinal Survey; (3) appropriate $200,000 annually to integrate the richness of the National Longitudinal Survey database into the ongoing functions of the BLS; (4) construct data tapes that contain the matched records of individuals using existing cross-sectioned files, thereby creating from existing information an entirely new source of longitudinal information on the labor market behavior of individuals; (5) develop methodologies to correct the biases inherent in the Gross Flow data, with the goal of making these data generally available; (6) create a panel of households that would be interviewed over a long-term basis using the monthly Current Population Survey instrument; (7) investigate the possibility of constructing an archive of unemployment insurance wage records; and (8) study the efficacy of conducting surveys that match both workers and firms. (25 references) (CML)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED317713

44. DATA NEEDS FOR LABOR MARKET ANALYSIS

Michael W. Horrigan

Commission Staff and
Bureau of Labor Statistics

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CE054129

Opinions stated in this document do not necessarily represent the official position or policy of the U.S. Department of Labor, Commission on Workforce Quality and Labor Market Efficiency.

44. DATA NEEDS FOR LABOR MARKET ANALYSIS

Michael W. Horrigan
Commission Staff and
Bureau of Labor Statistics

Any examination of data needs for labor market analysis requires taking into account a variety of considerations:

* Who are the users of existing data?

* Do current data sets adequately fit user needs? In particular, to what extent have the methodological approaches adopted by data users been compromised by the nature of existing data?

* How can existing data sets be revised or otherwise improved to take problems of method into account? A related point is whether or not existing data sets are being used to their full potential. This may involve issues of data accessibility or employing advanced computer technology to link files in new and useful ways.

* What are the lines of inquiry for which empirical estimates cannot be generated owing to a total lack of data; a lack of data of sufficient quality; or a lack of data which appropriately measures the relevant concepts?

Although these considerations indicate an ambitious agenda, the goals of this paper are more modest and realistic. Rather than develop a broad theoretical model which attempts to capture all of the above considerations, the design of this paper is to begin with a view of specific data needs and use the above considerations to justify the accompanying recommendations.

In developing the recommendations which follow, opinions were solicited from a wide spectrum of data users including government officials and the authors of our commissioned papers. Another source of information were the numerous reports issued by various intergovernmental agency groups, task forces, and other organizations

which have dealt with specific data issues. The general conclusions which have emerged from this effort can be summarized as follows:

1. The Department of Labor needs to improve its ability to generate timely labor market information on policy issues.

2. Labor market information with demographic and socioeconomic detail on individuals, especially longitudinal panel data, is invaluable to both researchers and policymakers and needs to be strongly supported.

3. There is a serious need to expand our base of knowledge on the role of firms in labor markets. While much useful information exists, there are significant gaps in our current data collection efforts.

4. There is a growing consensus that certain questions such as the extent and quality of worker training can only be effectively addressed through data sets which contain information on both firms and workers within those firms. Research on the viability and cost effectiveness of such data collection efforts needs to be conducted.

5. There is a significant commonality of interest existing in the world of data collection, presentation and analysis--ties that bring together the interests of data users and providers alike. Issues such as the latest developments in longitudinal analysis or the tradeoffs between the development of micro data files and confidentiality can be usefully explored in a forum bringing together the relevant users and providers of labor market data.

The recommendations given next follow this broad framework. The discussion justifying each recommendation builds upon a variety of concepts. Two which are useful to mention at this point are part of the nomenclature used to describe the characteristics of data sets--the sampling frame and the sampling view. The former determines the relevant population under discussion. For example, the Current Population Survey (CPS) is a nationally representative sample of individuals whereas the Longitudinal Establishment Data file is a representative sample of firms in the manufacturing sector.

A potentially important and distinguishing feature of sampling frames is whether or not the survey depends on the recall of individuals--asking respondents to recall or estimate a concept to the best of their knowledge--or if the survey relies on the establishment records of a firm. The former is usually a characteristic of surveys of households; the latter of establishment surveys.

The sampling view determines how the data are collected and presented. For example, the data could represent a summary of activity at a point in time--that is, a cross-sectional snapshot; or it could be a collection of retrospective information collected at a survey date--such as the information on work experience the previous calendar year collected in the annual supplement to the March CPS; or possibly a panel view in which the same individuals are reinterviewed over an extended period of time--that is, a longitudinal data set. In the discussion which follows these terms associated with sampling frames and sampling views are used frequently.

I. The Need to Inform Policymakers in a Timely Fashion

Some of the most frequent users of labor market information are agencies within the government, including the Council of Economic Advisors to the President, Congressional committees, as well as policy staffs located in the various agencies of government. Their needs range from long-range evaluation of government interventions in the labor market to much more short-run concerns such as obtaining timely information on current trends in labor markets. While much of this report points to the need to improve the availability and quality of

data for long-term research and policy analysis, there also exists an important gap in our ability to generate timely cross-sectional information on a wide variety of specialized labor market issues.

This judgement is not meant to imply there is currently no such capability. On the contrary, the Bureau of Labor Statistics has a proud tradition of being able to provide accurate information, rich in detail, on a highly diverse set of issues in a quick and timely fashion. In particular, the Bureau has set up a sampling frame of establishments permitting quick response surveys be taken of firms. In the past two years, this has included, among others, surveys on child care practices and drug testing programs in the workplace.

However, the ability of the Bureau to respond quickly to requests which require information on households can be limited. The Bureau's source of information is the Current Population Survey (CPS) which is conducted by the Bureau of the Census and is based on a sampling frame of household addresses drawn from the Decennial Census. Unless the information can be inferred from answers to the regular battery of questions in the monthly CPS, the Bureau must contract with the Census Bureau to conduct a supplement to the survey. The amount of time required to include the relevant questions on supplements to the survey, collect, process and analyze the data can be very significant.

As an example, the subject of worker dislocation became a major concern of policymakers during the 1981-82 recession. In particular, as a result of language in the October 1982 Job Training and Partnership Act, the Employment and Training Administration of the Department of Labor requested that the Bureau of Labor Statistics identify the

population of dislocated workers as well as their reason for dislocation, industry affiliation, and numerous demographic characteristics. However, because the dislocated worker concept was not currently identified by the monthly CPS, it was necessary for the Bureau to use a supplement to the CPS to gather the available information. The time costs of negotiating with the Census Bureau as to the timing of the supplement, creation and approval of the questionnaire, conducting the survey, processing and analyzing the data, resulted in a release of the first report on dislocated workers in November 1984.

Recommendation 1. Creation of a quick response household survey capability at the Bureau of Labor Statistics would add needed flexibility and timeliness to its ability to answer requests for data. Such a quick response survey would either use a random digit dialing method to create the sampling frame, supplemented by a purchase list for households without telephones or a list area frame which would also have uses for numerous other BLS programs. The organizational and budgetary structure would permit 2 surveys to be conducted a year at an annual cost of around $750,000 to $1 million dollars in 1988 dollars (all dollar estimates are given in 1988 dollars in this report). With the same amount levied for an initial appropriation, it would be possible to cover all startup costs and conduct one survey in the first year of operation.

II. Longitudinal Data Needs

Progress in the fields of labor economics, sociology, demography, and health economics, among others, has made significant advances owing to the development and availability of longitudinal data. As these data have developed, new approaches to empirical questions have come under the scrutiny of the social scientist. In particular, the perspective provided by cross-sectional data--stratified measures of the characteristics of the population at a point in time--are being

reexamined with the ability to follow the experiences of individuals over extended periods of time.

One telling example is the estimation of the length of time individuals remain unemployed. Reliance on cross-sectional estimates produces biased measures owing in part to the fact that such data capture individuals while their unemployment is still in progress--without providing information on the eventual length of spells. Longitudinal data hold the promise of following individuals from the time they enter unemployment until the date they either find employment or withdraw from the labor force. As a result, analysts are better able to capture the effects of changing economic and institutional conditions, such as the impact of recessions or the provision of extended benefits on the length of time people remain unemployed.

The previous example is but one of many which could be chosen to demonstrate the value of longitudinal data. A brief and partial listing of other important topics includes:

* retirement decisions and the effect of changing pension and social security rules,

* long-term earnings and employment effects of education and training programs,

* fertility and child-bearing decisions, especially the problems of adolescent mothers,

* experiences of veterans and the relationship between military training and subsequent career decisions,

* impact of social welfare programs on welfare dependence, earnings independence, and family structure.

Before discussing specific sources of longitudinal data, it is useful to note some of the special problems these data often pose. One obvious one is the difficulty in tracking individuals over an extended period of time. The longer you attempt to follow a cohort, the higher is the non-response and the refusal rates. As a result, more resources have to be devoted to the survey and interviewers have to work harder to preserve the integrity of the survey. One particular problem in this regard is the need to follow individuals who move.

Finally, there is increasing concern about data linkages. As longitudinal data sources proliferate, it is important that any proposal to match individuals between different data files protect the privacy of the individual and preserve the trust of the individual respondent. If not, the integrity of entire surveys can be called into question.

A. The National Longitudinal Survey

One especially valuable source of longitudinal information is the National Longitudinal Survey (NLS) which for the last three years has operated under the direction of the Bureau of Labor Statistics and for the last twenty years has been administered by the Center for Human Resources at Ohio State. There are three currently active cohorts: Mature Women (aged 30-44 in 1967); Youth (aged 14-21 in 1979); and Young Women (aged 14-24 in 1968). Two cohorts were discontinued in the early 1980's because of budget constraints; Older Men (aged 49-59 in 1966); and Young Men (aged 14-24 in 1966).

During the last two years, the Bureau of Labor Statistics has reviewed the use of the NLS program, and as part of this review process

issued a report on the future of the NLS released January 20, 1989. Among the recommendations made in the report, the Bureau has identified the following which are consistent with and speak directly to the broad mandate of the Commission:

Recommendation 2. The review of the NLS program has given strong priority to starting a new youth cohort. The data which has been collected to date from the cohort aged 14-21 in 1979 has proven extremely important in understanding the process of transition from adolescence to adulthood. The ability of agencies outside the Department of Labor to add supplementary sections to the core interview has been very cost effective, especially relative to the cost of mounting an independent survey effort.

The creation of a new youth cohort will permit examination of numerous issues identified as especially important by this Commission, including, but not limited to:

* the transition from school to work, including the problems of high school dropouts,

* the impact of training programs on the employment stability and wage profiles as individuals in the cohort progress into adulthood,

* the effect of various levels of educational attainment on labor market outcomes.

The cost of mounting a new survey effort is estimated to be about $4 million per year. A new youth cohort should be interviewed yearly for the first ten years, and then placed on a bi-annual interview schedule.

One of the problems affecting many longitudinal surveys, especially the NLS, is the effect of budget constraints on the ability to continue a consistent schedule of reinterviews for each cohort. At

the same time, one of the possible benefits of a stringent budgetary climate is the elimination of unnecessary duplication of effort across surveys.

However, the general funding situation for the NLS has been so tight in the last few years that the BLS faces and has already made tough choices between highly valued aspects of the survey. For example, one important aspect of any survey, data accessibility, has not been developed to its potential owing to a general lack of funds. As a consequence, the integration of the survey into the highly valued system of preparing data analysis for the public and for policymakers has not been accomplished. Instead, certain budget choices, such as putting the 1979 youth cohort on a bi-annual interview basis, are being considered to free up resources to permit continuation of other cohort interview schedules.

Recommendation 3. We recommend an appropriation of $200,000 a year for the purpose of integrating the richness of the NLS data base into the ongoing functions of the Bureau. These monies will pay for the added personnel and computer resources needed for this effort.

B. The Current Population Survey

Although a cross-sectional survey by design, the Current Population Survey offers a rich source of hitherto underutilized longitudinal data. The survey which is conducted by the Bureau of the Census under contract to the BLS, is based on a sampling frame of household addresses drawn from the Decennial Census. A household agreeing to participate in the survey will be interviewed for four

consecutive months, remain out of the survey for the next eight months, and be subsequently interviewed for another four periods. As a result, it is possible to systematically construct numerous longitudinal profiles across the eight possible months of information gathered on the individuals in each household. Although such matches are done on an ad hoc basis in response to requests from data users, budget constraints have prevented taking a systematic and integrated processing approach.

Recommendation 4. We recommend the construction of data tapes which contain the matched records of individuals using existing cross-sectional files. If this is done, an entirely new source of longitudinal information on the labor market behavior of individuals would be constructed entirely from existing information. The main costs will be computer processing and data verification expenses which, given the technical expertise of the Bureau, promises to be cost effective.

Another data set that can be derived from CPS cross-sectional files are the set of month-to-month transitions of individuals between employment, unemployment, and being outside the labor force--often referred to as Gross Flow data. By matching individuals between two consecutive survey periods, it is possible to identify the flows occurring between these labor market states. Given its national representativeness and the richness of demographic detail on individuals, researchers have long sought to exploit the potential of this data base. However, a problem exists in the data which, until resolved, prevents the realization of this potential.

As an illustration of the problem, consider the measurement of a change in the level of employment between two consecutive survey periods. By design, this change should equal the level of employment in

the first month adjusted positively for the flows into employment and negatively for the flows out of employment during the month between the survey dates. That this simple accounting scheme is not satisfied is a problem that has prevented more proactive use of these data. There are, however, promising statistical avenues which can be pursued for increasing the reliability of these data.

Recommendation 5. Resources should be provided for the development of methodologies to correct the biases inherent in the Gross Flow data with the eventual goal of the release and general availability of these data.

One feature of the Current Population Survey which distinguishes it from other surveys is its sampling frequency. By matching the monthly responses of individuals during their time in the survey, one can build a longitudinal profile that is relatively less prone to recall error than surveys with reinterviews that take place every 4 months, or each year, or even every two years. This unique feature of the survey permits a greater precision in describing the month to month activity of respondents. The drawback of the survey's design in this regard is that individuals are in the survey for a maximum of four consecutive months. However, given that the structure for drawing the sampling frame is already in place, it is possible to enlarge the sample at costs which are much lower than beginning a new survey. Accordingly, it is economically feasible to create a new longitudinal sample using the monthly CPS sampling universe and sampling frame.

Recommendation 6. Resources should be provided for the creation of panel of households which would be interviewed over a long-term basis using the monthly CPS survey instrument. One critical feature of this

survey which is not part of the current sample design of the CPS, is the need to follow individuals in the sample who move out of the household being surveyed.

C. Retirement History Survey

In May 1988 the Ad Hoc Advisory Panel to the Behavioral and Social Research Program of the National Institute on Aging (NIA) issued a set of recommendations on priorities for data collection in Health and Retirement Economics. The report concludes:

> "...available datasets pertinent to health and retirement economics are widely perceived to be inadequate to address current and future concerns about the aged."[1]

There are numerous areas of policy which, in the opinion of the Panel, can benefit from quality sources of data on aging. One of the most cost effective benefits would be the ability to inform policy on health expenditures--especially Medicare and Medicaid; labor force participation and retirement decisions of the elderly; housing and living arrangements--especially their relationship to long-term care; and the general financial status of the elderly.

The key to understanding the issues underlying the development of data on the elderly centers on the relationship between two alternative proposals which emerged from the Panel's report. The first is the development of a new Retirement History Survey. This option is based in part on the premise that "a wholly new data collection effort would not have to make concessions to history and could get it right the first time."[2]

The second proposal is to add question modules to existing longitudinal surveys to reflect the interests and needs of the community conducting research on aging. The Panel purposefully left this option open, relegating the details of its technical feasibility to a chapter discussing each of the relevant existing data bases. A reading of the Panel's report indicates that the most favored candidate for inclusion in this strategy is the National Longitudinal Survey (NLS). However, no specific recommendations were made.

One additional complication to the current deliberations is the impact of Title 13 privacy regulations. The Census Bureau is in the early stages of planning a new retirement history survey. The Panel offers the opinion that "an integral part of any data collection strategy is that the resulting data be easily and widely available to the research community."[3] The Panel report goes on to conclude, "It would behoove the NIA to await an official Census Bureau policy on access before funding any data gathering that would fall under Title 13 regulations."[4] This effectively eliminates from current consideration both the Census retirement survey effort as well as a reinterview of the Title 13 governed original retirement history survey.

Recommendation Options. What conclusions can be reached from these deliberations? First, we concur with the Panel's report on the importance of continued support for data collection on issues of aging. Second, one course is to recommend support for the Panel's second option, supplementing the National Longitudinal Survey with question modules designed to capture needed information on aging. There are two directions this recommendation could take--either supplementing the Mature Women's Cohort (aged 30-44 in 1967), or reinterviewing the Older Men's cohort (aged 45-59 in 1966).

A bolder recommendation is to offer support for the Panel's first option, a new retirement history survey. If the Commission decides to move in this direction, the Commission should be cautious about one additional option the National Institute on Aging is considering; namely, the option of using existing sampling frames from the National Health Interview Survey (NHIS) or the Panel Survey of Income Dynamics (PSID) to develop the new retirement history sample. In particular, there are some significant concerns that using an existing sampling of individuals to form the basis for a new survey would violate the spirit of trust between the data provider and the individuals in the survey and, as a result, threaten the integrity of the effort.

III. Measurement of Labor Turnover and Labor Shortages

As surprising it may seem to some, there does not currently exist direct and accurate measures of either job vacancies or labor turnover in the United States. With regard to the former, then Secretary of Labor Ann McLaughlin made the observation that:

> "The Bureau of Labor Statistics has examined available measures of labor shortages--by industry, occupation, and geographic area. Each of these measures is indirect...Unfortunately, the current labor market information system contains neither data on job vacancies nor data about the time it takes employers to fill specific jobs...Also useful would be a direct measure of new hires in order to better analyze job vacancy data with respect to labor shortages."[5]

Information on job vacancies could provide needed measures of how demand for labor changes with the business cycle and with the relative tightness of local labor markets. Such information would be invaluable to researchers. As well, policymakers could benefit from the ability to identify geographical, industrial, and occupational areas in which there are shortages of workers. Such information may help determine areas in which we are experiencing a growing skills gap; it may also help to

carry out the requirements of the Job Training Partnership Act (JTPA) program, and labor certification of aliens, to name a few.

The measurement of job vacancies is an attempt to capture the dynamic nature of labor markets. Vacancies are, however, but one important aspect of these dynamics. The other is the concept of labor turnover. Each month the CPS produces measures of the levels of employment and unemployment as well as the number of individuals not in the labor force. From one survey month to the next, the level of unemployment changes due to additions to the unemployment pool in the form of quits, layoffs, fires, retirement and death, and deletions from unemployment in the form of withdrawals from the labor force, new hires and rehires. The flows that are responsible for the changes in the stock levels of employment can be sizeable and very responsive to changing business cycle conditions.

For example, in January 1980, the peak month of the 1975-1980 economic recovery, the level of unemployment stood at 6.8 million individuals. Underlying this change in the stock level of unemployment, however, was a highly dynamic process. Using Gross Flow data from the CPS, an estimated 49 percent of these individuals left unemployment by the next month. As well, between the January and February surveys, 3.1 million new individuals became unemployed, representing 46 percent of the January unemployment level.

These numbers only reflect the overall levels of additions and deletions. Currently there does not exist a source of data on the individual components of labor turnover in the U.S. economy. Many may recall the labor turnover series which was discontinued as a result of

budget cuts in 1982. Although informative, those data were only representative of labor force movements in and out of the manufacturing sector. Given the substantial shifts which have occurred in the recent years between manufacturing and the service sector, it is imperative that labor turnover information be representative of all industries.

As well, some may recall a short-lived vacancy series which ran between 1969 and 1972, also limited to the manufacturing sector. Since the discontinuation of both series, the processing environment for conducting establishment surveys has improved tremendously, resulting in more efficient data transmittal mechanisms (especially through computer assisted telephone interviews and voice recognition mediums), substantially smaller response burdens on the part of firms, and generally more reliable data. However, it is paramount that any effort to measure job vacancies and labor turnover be done carefully and thoroughly from both a cost perspective as well as reflecting concern for data quality and respondent burden.

Recommendation 7. We recommend conducting a single pilot establishment survey which will permit separate judgement as to the efficacy of measuring job vacancies and the components of labor turnover. There are numerous cost savings to using a single pilot to evaluate two potential programs--but such a recommendation is not meant to imply that a vacancy series and a turnover series must co-exist.

In addition to measuring the levels of vacancies and the various types of labor turnover, this pilot would also examine the efficacy of collecting information on the occupational categories and geographical location of vacancies; the posted wage rates associated with vacant

positions; and study the respondent burden of providing wage information on new hires. The cost of setting up and conducting a pilot survey is estimated to be $350,000.

IV. Data from the Records of Establishments

There is a growing perception that in their current form, establishment data are not being used to their full potential, and there are significant gaps in our information on firms not being filled by existing surveys. One example has already been touched upon in this paper, the need for data on job vacancies, new hires and sources of labor turnover.

The recommendations given in this section and the sections that follow outline the need for systematic research on the types of information collected from firms. Although the case for collecting more data from establishments is a compelling one, we recommend that a cautious approach be taken. One reason for this position is the need to protect the confidentiality of firms. The continuing call for very detailed information on firms has to be balanced against the need to protect the identity of firms in the data base. Moreover, since much existing data on firms is collected through the system of State Employment Security Agencies, any effort to create national data bases from their files must respect the integrity and autonomy of these organizations.

One particular example reflecting these concerns is the idea of creating a national archive of the wage records firms provide to State agencies under the Unemployment Insurance (UI) system. Combining these

data into a national data base offers the potential of a rich new source of information but it also raises the issues of data confidentiality and the need to carefully structure cooperation between State agencies. It is by no means clear that an effort to combine and archive state UI records would be successful, but the potential such a data set represents suggests that a carefully designed pilot study be constructed.

Through the auspices of the U.S. Department of Labor, the Northeast-Midwest Institute prepared a report released on January 10, 1989 which examined the feasibility and desirability of creating a longitudinal data set based on wage records collected for the administration of the Unemployment Insurance System. The report concluded that owing to the extensive coverage of the reporting system (nearly 97 percent of total employment), near uniformity of reporting practices across states, and rapid advances in computer technology, the creation of a longitudinal data set of individual wage records using establishment data reported by firms under the UI system is technically feasible and economical.

The desirability of such a data set can be demonstrated on several grounds. In particular, there are two significant constituencies of users: academic researchers and policymakers. On the research side, examples of issues include:

1. A greater understanding of the mobility patterns of workers in terms of industrial affiliation and geographic location.

2. The need for information characterizing the adjustment of immigrants to labor market.

Besides the benefit to the research community, nationally archived UI data would be an invaluable tool in the evaluation of Federal and State expenditure programs. In general, the impact of programs in terms of their employment, earnings, and worker dislocation effects could be examined. More specifically, a nationally archived UI data base offers a reliable source of information for the evaluation and the establishment of performance standards for government programs such as JTPA.

Already, a number of states have begun using UI claimant data to facilitate their evaluation studies of the JTPA program. A national longitudinal data base would improve this capability, especially in terms of following individuals who move across state boundaries after receiving training. Moreover, the identification and measurement of the earnings of proper control groups is found to be much more reliable, less expensive and less prone to statistical problems--such as selection bias--than conventional methods which rely on household data.

There are numerous other programs where evaluation is based in part on labor market indicators such as employment stability and earnings profiles, and as a result, evaluation efforts could benefit substantially from a longitudinal UI data base. For example, in the area of income maintenance programs, by matching longitudinal UI records to welfare caseload files, the success of trying to reduce welfare dependency could be examined.

Recommendation 8. Resources should be provided for the setup of a pilot project to systematically investigate the possibility of constructing an archive of UI wage records. The project should be conducted on a regional basis, and besides investigating the costs of creating such a

data base, careful consideration should be given to issues of confidentiality and preservation of the integrity of State Employment Service programs.

V. The Feasibility and Desirability of Surveys which Match Workers and Firms

During the first meeting of the Commissioners on September 19, 1988, Commissioner Gary Becker decried the lack of reliable data on the nature and extent of training which takes place in firms--or sponsored by firms. In attempting to address Commissioner Becker's request that the reasons for such a lack of data be examined, one is immediately struck at the difficulty of the task.

Using the experience of previous attempts to measure training as a guide, one finds most surveys have significant weaknesses in their overall concept and design. First, household surveys suffer from the drawback that respondents often do not know the funding auspices under which they received their training, how many individuals are trained each year, the success rate, and so on.

The strength of household surveys is that they do allow researchers to get the respondent's description of his or her previous training as well as the nature of the most recent training episode. Such baseline information is useful in evaluating the extent to which the current training has been effective. A problem in this regard is that most household surveys have been cross-sectional, and although some retrospective information is collected, there have been few attempts to collect longitudinal profiles of the effectiveness of training over an extended period of time. The one significant exception are the training

questions asked in the National Longitudinal Survey. These questions are probably the best source of information on training available.

Second, an establishment survey designed to collect data on the provision of training would not by itself be very informative. Firms can provide information on the nature and number of formal training opportunities offered to workers--but information on the characteristics, abilities and reactions of the participants would probably be lacking.

An obvious approach would be to use a combined household and establishment survey approach. One immediate question is whether or not the survey should be conducted as a random sample of establishments, followed by a random selection of workers in the establishment--or should the survey be conducted as a random survey of individuals with follow-up contacts of firms.

The proper sampling frame is a question for which no well-defined answer currently exists, pointing to the need for research. However, in considering the overall idea of matching workers and firms, one begins to see the potential for analysis of areas beyond the examination of training issues. In particular, there is a growing call for data which relates organizational aspects of the firm to information on the characteristics of workers. In a 1986 report entitled "America at Work: National Surveys of Employees and Employers," the Social Science Research Council reported to the Department of Labor:

> "There is, however, a set of issues that escapes attention in (labor market) surveys (or is at best marginally treated): the impacts of different organizational conditions of work on worker productivity, worker well-being and satisfaction, labor force

behaviors such absenteeism and quitting work, and labor-management and industrial relations."[6]

One of the principle pieces of information needed to evaluate worker/firm surveys are their cost. In addition, there are also considerations of confidentiality as there are with any survey design. However, given the potential this approach represents, it seems desirable to devote resources to a systematic study of these questions.

Recommendation 9. Resources should be provided for the study of the efficacy of conducting surveys which match both workers and firms. Given its experience with both household and establishment surveys, the Bureau of Labor Statistics should be given responsibility for studying this issue.

The Bureau already has a universe file set up for drawing a sampling frame of establishments. Moreover, BLS experience with establishment data would permit a careful study of the issues of cost and confidentiality requirements associated with a study which matches firm and worker data. Another natural starting point is to use the National Longitudinal Survey to examine the viability of starting with a household based survey and collecting information from firms, especially with respect to training.

V. Communication on Labor Market Data Issues

Progress in the world of data collection and dissemination has been substantial in the last twenty or so years. This progress comes from both technical advances in the pure quality of the data as well as in the usefulness of data to policymakers, researchers and the general

community of data users. Indeed much of the focus of the recommendations thus far has been on ways to improve on what are generally perceived as highly valued sources of information.

In attempting to review data needs for labor market analysis it is useful to keep in mind the following three important observations: First, the worlds of both data providers and users are large and extremely diverse. Second, despite this diversity there are a number of issues which reflect a commonality of interest between these diverse groups. Third, in general, both data providers and data users have established effective formal mechanisms of soliciting advice and opinions from the world of data users. Examples include the Advisory Committees of the Bureau of the Census, the Council of Professional Association of Federal Statistics (COPAFS), the Association for Public Data Users (APDU), and the Bureau of Labor Statistic's Labor Research Advisory Council (LRAC), and Business Research Advisory Council (BRAC). Sometimes, the advice is informal, depending on the nature of relationships established between the data provider and the entire data user community.

In conducting this overview of data needs for labor market analysis it also became apparent that there are significant umbrella issues of particular relevance to a wide spectrum of data users and providers. For example, one topic which has emerged as very important in our discussion of data needs are proposals to improve the accessibility and close the information gaps in establishment data. Another are issues relating to the development, quality, and coverage of longitudinal data bases. A third are general issues relating to

confidentiality, statistical measures of data quality, and appropriateness of survey measures of theoretical concepts. A fourth are issues related to the integration of advances in computer technology into the way in which data are released.

The opportunities for diverse interests to come together and discuss these topics of mutual interest are somewhat limited. Even though professional meetings such as the Annual Research Conference at the Bureau of the Census are of significant value, they do not necessarily reflect broader issues related to labor market data. What is needed is a flexible structure in which the participants and the aspect of labor market data being considered can change from one meeting to the next.

Recommendation 10. Resources should be provided to set up an annual conference of data providers and data users. By design these conferences should be organized around a single theme each year, and the participants of the conference should reflect this design. To achieve continuity, the conference should be sponsored by and organized under the auspices of the Department of Labor. The invitees should be drawn from the various constituencies of data providers and users relevant to the discussion.

NOTES

1. Report of the Ad Hoc Advisory Panel to the Behavioral and Social Research Program. *National Institute on Aging Recommendations to the NIA Extramural Program on Priorities for Data Collection in Health and Retirement Economics* (May 1988): p. iii.

2. ibid, p. vi.

3. ibid, p. vii.

4. ibid, p. vii.

5. Report of the Secretary of Labor, *Labor Market Shortages* U.S. Department of Labor (January 1989): p. 20.

6. Report of the Social Science Research Council Advisory Group on a 1986 Quality of Employment Survey. *America at Work: National Surveys of Employees and Employers* (April 14, 1986): p. 7.

REFERENCES

Bishop, John. "Policy Evaluation with Archived Wage Record Data: Limitations of Existing Data Sets." *Invitational Meeting of the Northeast-Midwest Institute* (January 10, 1989): 13-24.

Bishop, John. "The Recognition and Reward of Employee Performance." *Journal of Labor Economics* vol. 5, no. 4, pt 2: S36-S56.

Bulletin 2285. *Handbook of Methods of the Bureau of Labor Statistics* U.S. Department of Labor, Bureau of Labor Statistics. (April 1988).

Bulletin 2310. *Permanent Mass Layoffs and Plant Closings, 1987* U.S. Department of Labor, Bureau of Labor Statistics (July 1988).

Cohen, Malcolm. "Developing a Wage Record Data Archive: Some Implementation Issues." *Invitational Meeting of the Northeast-Midwest Institute* (January 10, 1989): 35-43.

Davis, Steve J. and Haltiwanger, John. "Gross Job Creation, Gross Job Destruction and Employment Reallocation." *Preliminary Draft for presentation at the 1988 Econometrics Meetings*, (December 28-30).

Duggan, Paula. *Labor Market Information: Book I, A State Policymaker's Guide* Northeast-Midwest Institute: The Center for Regional Policy (1988).

Employment Service-202 Operating Manual (Part III., Sections 0400-0599, Employment Wages and Contributions) U.S. Department of Labor, Bureau of Labor Statistics

Hamermesh, Daniel S. "Data Difficulties in Labor Economics." *Fifty Years of Economic Measurement*, Ernst Berndt, Erwin Diewert and Jack Triplett (editors), University of Chicago Press, (1989), forthcoming.

Hanna, James. "The State Perspective on a National Wage Record Database." *Invitational Meeting of the Northeast-Midwest Institute* (January 10, 1989): 25-34.

Jacobson, Louis and Thies, Carolyn. "Updating the Pennsylvania 5-Percent Work-Claim History File to Cover 1969-85." *Final Report to the Bureau of International Labor* U.S. Department of Labor (August 1, 1988).

King, Christopher. "Employment and Earnings Dynamics in the 1990's: Policy Issues and Potential Uses of National-Archived UI Wage Records." *Invitational Meeting of the Northeast-Midwest Institute* (January 10, 1989): 1-12.

McGuckin, Robert H. and Pascoe, George A., Jr. "The Longitudinal Research Database (LRD): Status and Research Possibilities." *Discussion*

Paper CES 88-2 Center for Economic Studies, Room 3442, FOB 3, U.S. Bureau of the Census, Washington, D.C. 20233. (July 1988).

National Commission Employment Unemployment Statistics. *Counting the Labor Force*. Appendix Volume II. (Labor Day, 1979).

National Commission Employment Unemployment Statistics. *Data Collection, Processing and Presentation: National and Local, Appendix, Volume II to Counting the Labor Force*. Appendix Volume II. (Labor Day, 1979).

Organization for Economic Co-operation and Development. "Chapter IV: The Process of Job Creation and Job Destruction." *Draft Report, Working Party on Employment* (April 29, 1987).

Report 718. *Major Programs of the Bureau of Labor Statistics* U.S. Department of Labor, Bureau of Labor Statistics (April 1985).

Report from the National Science Foundation Conference on the Future of the National Longitudinal Survey and the National Longitudinal Survey Technical Review Committee. *The Future of the National Longitudinal Survey*. Center for Human Resource Research at the Ohio State University. (January 20, 1989).

Report of the Ad Hoc Advisory Panel to the Behavioral and Social Research Program, National Institute on Aging. *Recommendations to the NIA Extramural Program on Priorities for Data Collection in Health and Retirement Economics* (May 1988).

Report of the Secretary of Labor. *Labor Market Shortages*. U.S. Department of Labor (January 1989).

Report of the Social Science Research Council Advisory Group on a 1986 Quality of Employment Survey. *America at Work: National Surveys of Employees and Employers* (April 14, 1986).

Stevens, David W. and Duggan, Paula. *Labor Market Information: Book II, An Agenda for Congress* Northeast-Midwest Institute: The Center for Regional Policy (1988).

User's Guide to Labor Force Statistics from the Bureau of Labor Statistics U.S. Department of Labor, Bureau of Labor Statistics, Office of Employment and Unemployment Statistics (1985).

Utter, Carol M. "Labor turnover in manufacturing: the survey in retrospect." *Monthly Labor Review* (June 1982): 15-17.

Waite, Linda J. "Sociologists' Need for Longitudinal Data: The NLS Cohorts" *Presented at the National Science Foundation Conference on the National Longitudinal Surveys*, Washington, D.C., (October 21, 1988).